

Forecasting Italian Electricity Zonal Prices with Exogenous Variables

Angelica Gianfreda¹ and Luigi Grossi²

Abstract

In the last few years we have observed deregulation in electricity markets and an increasing interest in price dynamics has been developed especially to consider all stylized facts shown by spot prices. Only few papers have considered the Italian Electricity Spot market since it has been deregulated recently. Therefore, this contribution is an investigation with emphasis on price dynamics accounting for technologies, market concentration, congestions and volumes. We aim to understand how these four variables affect zonal prices since these ones combine to bring about the single national price (*prezzo unico d'acquisto*, PUN). Hence, understanding its features is important for drawing policy indications referred to production planning and selection of generation sources, pricing and risk-hedging problems, monitoring of market power positions and finally to motivate investment strategies in new power plants and grid interconnections. Implementing Reg-ARFIMA-GARCH models, we assess the forecasting performance of selected models showing that they perform better when these factors are considered.

Key words: Electricity prices, Production technologies, Market power, RSI, Congestions, Volumes, Fractional Integration, Forecasting, Policy

1. Introduction

Several empirical features of electricity prices observed at daily frequency have been widely discussed: mean-reversion, seasonality, time varying and

¹Jean Monnet Fellow at the Robert Schuman Centre for Advanced Studies (RSCAS), European University Institute (EUI). Corresponding author's email angelicagianfreda@yahoo.it.

²Department of Economics, University of Verona. Email: luigi.grossi@univr.it

clustered volatility, inverse leverage effect and extreme values called spikes or jumps, see for instance Escribano et.al. [13], Knittel and Roberts [38], Koopman et al. [39] and Gianfreda and Bunn [19] among others. While seasonality and clustered volatility are well-known features, the remaining stylized facts require to be better explained. Mean-reversion is the tendency that prices show tending to a long-run mean level. The inverse leverage effect, discovered by Knittel and Roberts [38], is the inverse reaction to shocks: electricity price volatility tends to rise in presence of positive shocks more than in presence of negative ones. Extreme values or spikes are results of abnormally large variations in price caused by weather conditions, outages or transmission failures.

A stylized fact that has been fairly neglected so far is the long memory of the process generating electricity prices. When unit root tests are applied usually the presence of unit root is rejected. On the other hand, stationarity tests provide evidence of non stationarity. Moreover, when the empirical correlation function is estimated and visualized through a correlogram, a long memory pattern can be observed because autocorrelation tends to decrease very slowly as the lag increases. These combined results lead to the conclusion that the analyzed time series could be generated by a fractionally integrated process. This feature is explored for the first time using data of the Italian market, whereas this methodology has been previously implemented by [29] and [39], for instances on Nordic, German, French and Dutch markets. Previous papers about the Italian Power Exchange (18; 48 and 5) have completely discarded this peculiarity and have focused on prices and returns.

Beside fractional integration, we considered also the conditional heteroscedasticity of residuals which has been captured by GARCH models. The final correlation structure of electricity prices has been then captured by Reg-ARFIMA-GARCH models. An interesting result is that the Italian market does not show the presence of leverage effect, nor direct or indirect, as explored by Gianfreda [18].

The procedure suggested in this paper is aimed to correctly identify the appropriate stochastic generating process for electricity prices which is important for several reasons. First of all, the price dynamics can be used to understand the deregulation process, verify the competition in this electricity market and give indications on spot and forward price definitions (21). Secondly, a good model identification leads to proper managing of network congestions for needs of continuous real time balancing. Thirdly, modeling

is important for forecasting, trading, generation planning and plants availability, for risk management and hedging purposes in such market given the launch of the Forward Electricity Market (MTE)³ on 3 November 2008.

Another original contribution of this work is the use of exogenous variables to explain the electricity price dynamics. Exploiting the massive information provided by the Italian independent system operator (*GME*, Gestore dei Mercati Energetici⁴), we have analyzed the effect of technologies, market power, network congestions and demand on prices using 1461 daily observations from 1st January 2005 to 31st December 2008, extracted from an hourly database⁵ with quantitative indications for all explanatory variables and all considered zones. It is well-known that electricity prices depend on prices of generation sources employed, however there is no evidence on the degree and sign of these influences. Considering *technologies*, we have used the *marginal technology index (MTI)* provided on zonal basis by the GME which indicates the technology fixing the price over each zone, as explained later in Section 3.1. We moreover control for the exercise of *market power* from the generation side using the *residual supply index (RSI)*, described in Section 3.2. Considering the third variable, we follow [29] in defining a *congestion* event every time we observed different zonal prices on contiguous zones and its construction is explained in Section 3.3. Finally, we have also considered the influence of *zonal demands* on prices.

Therefore we aim to provide answers based on empirical evidence on how generation sources, market power, congestions and demand affect zonal

³All the abbreviations refer to the Italian names and definitions.

⁴The web-site is available at <http://www.mercatoelettrico.org>

⁵We decided to convert hourly data to daily data, as suggested in the literature subsequent to Lucia and Schwartz (2001) in view of the interpretation of our results. Policy indications can be drawn regarding the keys used in the paper: technologies, congestions and market power. In details, if investing in new under-utilized technologies (as for example wind and other renewables never found to be important in determining the zonal prices), the reader should keep in mind that new plants (or additional capacity) are realized considering daily wholesale prices corresponding to daily interest rates on funds necessary to undertake the investments, often evaluated according to the notion of the net present value - and this actually applies to both new generation as well as to network grid extension. In addition, several kinds of costs do affect economic evaluation of investments; and prices of fundamentals, as natural gas, oil, coal and CO₂, do represent costs with daily quotations for energy firms (we refer to Datastream, Bloomberg, Platts and Heren, among other data providers). Therefore, in view of our previous considerations, we have decided to adopt a daily horizon.

prices. Having a clear picture of these relationships, then it would be possible to obtain policy indications for future investments on an optimal technology mix, in additional capacity and in network interconnections. A procedure has been followed to select the variables which most significantly influence prices. Finally, combined models have been estimated for each zone in which the Italian market is organized.

The last contribution of the paper is the use of models with explanatory variables for short term forecasting. A rolling window procedure has been applied to assess the forecasting performance of the best model for each zone, analyzing the superiority of selected models with respect to the simplest ones. Furthermore, real out-of-sample forecasts have been obtained by means of a scenario analysis.

The paper is structured as follows: Section 2 links our research to the existing literature. The Italian zonal structure is explained in Section 3, where technologies, market concentration and congestions are also introduced and defined. Model specifications and results are studied in Section 4, whereas the forecasting performance is presented in Section 5. Section 6 concludes.

2. Background and literature review

Earlier contributions proposed several specifications for the electricity price process, taking into account traded volume, as in Goto and Karolyi [27], or price volatility, demand and margin as in Karakatsani and Bunn [37] and again power consumption and water supply as in Koopman et al. [39]. Hence we have found precedents, but none of these has been employed in the first empirical investigations on the Italian market, to the authors' knowledge. In addition, we detect important features of Italian spot prices implementing models with daily median prices accounting for spiky behavior, technologies determining zonal prices, indicators of market concentration, congestions among contiguous zones, and finally demand. Following Haldrup and Nielsen [29], we propose to consider possible congestions among zones, where a *congestion* is identified every time we observed different zonal prices. The technical factors underlying transmission network congestions may have a crucial influence over the behaviour of generators resulting in the allocation of production and this may affect the final prices paid for electricity. Hence generation, congestions and market power are strongly interdependent factors as in Furió and Lucia [16]. Therefore as Zarnikau and Lam [57] and Lisea et al. [41] point out, the transmission capacity plays an important role in

controlling congestions, reducing the impact of market power and improving market competitiveness.

In simple words, a generator has market power if it is able to raise the electricity price above marginal cost without experiencing a significant decline in demand. Previous studies focussed on this topic in the electricity generation sector relying on oligopoly theory, implementing simulation techniques to model the electricity generators' behaviour, see Green and Newbery (1992), Newbery (1998), and Wolfram (1998, 1999). Some others proposed empirical research as Wolak and Patrick (1997), Wolak (2000), and Borenstein et al. (2000), Helman [31], Bask and Widerberg [2]. For a survey on models to detect market power see Fridolfsson and Tangeras [15].

Traditionally, analysts and anti-trust regulators investigate market power issues using various measures of market concentration such as the popular *Hirschmann-Herfindahl* index (HHI), computed as the sum of the shares of the volumes sold in the market by market participants (see 46 and 4 among others). Market power can also be measured by the *Residual Supply* index (RSI), which gives indications on the presence of residual market participants necessary to cover demand. Since there is not a consensus on which measure is the best indicator of market power for the electricity markets, because there is a number of factors to account for (transmission constraints are an example), we have initially consider both structural indexes. However, since our results confirmed the common belief that HHI is not a good measure of market power, we have included only the second index in the final analysis.

Moreover, we address the issue of forecasting electricity prices since market participants need specific information on a short-term period to set their optimal bidding strategies, or on a longer term to base bilateral contracts. Therefore price forecasting is essential to both agents and practitioners.

As [37] and [53] suggest, it is possible to move from classical methods for the analysis of time series to models for unobserved components, considering dynamic regressions (as in 44 and 37), structural time series and ARIMA models (as 8; 7; 6; 20), jump diffusion (see 35; 50; 38) and regime-switching (as in 33; 55 and 30), among others techniques. However only few papers consider fundamental drivers or explanatory variables in assessing the forecasting performance. In details, demand, margin and scarcity were implemented in [37]; power consumption and water supply as in Koopman et al. [39]; whereas load and air temperature were used in [56]. In this paper, we instead consider new variables as production technologies, concentration and congestions in assessing the forecasting performance of selected models

for zonal prices.

3. The Italian Zonal Market

The Italian wholesale electricity market, or Ipex (*Italian Power Exchange*), started its operations in April 2004 but became an Exchange only in 2005 registering an increasing in traded volumes from 73 TWh in 2004 to 232 TWh in 2008. As other electricity markets, *Ipex* consists of the Spot Electricity Market (MPE), which is a platform for physical delivery; the Forward Electricity Market (MTE), launched on November 2008; and the platform for physical delivery of financial contracts (CDE).

Considering market mechanisms, the MPE consists of the Day-Ahead Market (*Mercato del Giorno Prima*, MGP), the Intra-Day Market (*Mercato di Aggiustamento*, MA or MI) and the Ancillary Services Market (*Mercato dei Servizi di Dispacciamento*, MSD). The Italian independent system operator, called *Gestore dei Mercati Energetici* (GME), operates on the day-ahead market (MGP) which is a single implicit-auction market, and not a continuous-trading market, pertaining to the next day. It opens at 8.00 of nine days before the delivery day, here participants start to submit their offers for sales and purchases until 9.00 of the day before the day of delivery, when the MGP closes. Then according to the economic merit order criterion and to the capacity limits of the transmission lines between zones, offers and bids can be accepted. The accepted supply offers are evaluated at the clearing price of the zone. This price is the equilibrium price determined on hourly basis by the intersection of the demand and supply curves. Hence the *zonal market clearing prices* are those prices observed on several zones or areas, and they can differ across zones if a proportion of the grid becomes congested and so separated from the entire network (54). These results are made known within 11.30 of the day before the day of delivery. On the other hand, the accepted demand bids, pertaining to consuming units, are evaluated at the single national price (*Prezzo Unico Nazionale*, PUN) which is the purchase price for end customers and it is computed as the average of the zonal prices weighted by zonal consumptions.

On the Intra-Day Market (MI, or adjustment market MA) opening at 10.30 of the day before delivery day, participants can modify their positions resulting from the MGP market by submitting additional supply offers and demand bids; but now the zonal prices are used to evaluate the accepted

purchase bids. MI takes place in four sessions (MI1, MI2, MI3 and MI4) with different opening and closing times⁶.

The transmission system operator, Terna S.p.A., starts its operations on the ancillary services market (*mercato del servizio di dispacciamento*, MSD) at 15.30, managing and controlling the power system, cross zonal congestions and real-time balancing. The MSD consists of an ex-ante MSD stage and of the Balancing Market (MB). Both markets take place in multiple sessions, according to specific dispatching rules. During the ex-ante MSD, Terna accepts demand bids and supply offers in order to provide reserve, relieve congestions and balance energy injections and withdrawals. It takes place in a single session on the day before the delivery day and the sitting for submissions of bids and offers opens at 15:30 and closes at 17:00, with results known within 21:00 of the day before the day of delivery.

The Balancing Market (MB) takes place in five sessions. The MB1 considers bids and offers submitted in the previous ex-ante MSD session. For the other sessions, all the sittings open at 23:00 of the day before the day of delivery (and not before the results of the previous ex-ante MSD session are known) and close one hour and a half before the first hour negotiated in each session. In the MB, Terna performs the real-time balancing.

Moving to the market structure, Italy is segmented into several zones as a consequence of congestions. The zones are classified in the following way: physical or national zones, virtual or foreign zones and finally limited production poles. Virtual or foreign zones are the neighboring markets connected to Italy: Corsica, France, Austria, Switzerland, Slovenia and Greece. The limited production poles only inject electricity into the systems, and among them we find Brindisi and Rossano. Finally, the national zones⁷ are: North, directly connected with Central North, which is at the same time indirectly connected⁸ with Sardinia and directly connected with Central South. This

⁶Details on market timing and mechanisms are available on [26].

⁷The market was born with this zonal structure in April 2004 and it remained unchanged until December 2008. After that, Calabria was included in the Southern zone, hence investigations refer only to a time period going from January 2005 to the end of 2008. This is to show that this market is comparatively young and there are continuous structural changes.

⁸This connection is called *indirect* because electricity flows through Corsica, which is a foreign zone not included in our analysis.

the previous ones. Hence the *marginal technology index* (MTI) represents these twelve technologies, which have been used in a previous investigation of Italian zonal price dynamics to detect influences of generation sources on price and volatility (see 20). Contrary to what done by the GME¹⁰, we have decided to cluster all previous technologies into the following six types of the MTI index for a better representation of zonal generations and distinguishing between oil, gas and coal producing plants¹¹: *Coal* (all multifuel and thermal power plants with coal), *Thermal* (plants without coal, or TNC), *Hydro*, *Wind* (renewables), *CC* that is combined cycles (CCGT and GT) and finally *Other* plants not included in the previous ones. As proposed in 2009, we compute for every group of technology the number of hours (frequency) in which it has fixed the price over the corresponding zone and we built a set of 6 dummies, one for each group, and we attributed one to the group with the maximum frequency over the day and zero to the others¹². Formally, let

¹⁰In the annual report 25 the following groups of technologies have been considered yearly and so at the national level: other, pumped storage, modulation, fluent, CCGT, thermal conventional, see page 99.

¹¹It is well documented that oil and gas have similar and correlated dynamics whereas coal has a dissimilar behavior.

¹²It must be emphasized that the solution adopted has a limitation because it specifies the major technology (or technologies with equal frequencies) setting the daily price in one zone. Therefore, it simply picks up the technology which was on the margin for a larger number of hours. And so, it is not able to capture a less marginal but more expensive technology with a stronger effect on the *average* daily price – that is another reason to justify our choice of daily *median* prices. In details, we have detected such cases in which we used *hydro*, and so a less expensive technology, whereas *Coal*, *TNC* and *CC* were important but less marginal. For *Hydro*, we observed 449 cases in North, 218 in CNorth, 183 in CSouth, 185 in South and 156 in Calb over a total of 1461 days; whereas for *Other*, we found 40 cases in North, 26 in CNorth and CSouth, 25 in South and 28 in Calb. Therefore, to account for this limitation, we computed the daily frequency each time that we underestimate the marginal technology costs with respect to a threshold equal to a third of a day (that is 0.333 given by 8 peak hours over 24). In simple words, every time we detected and used a cheap technology (mainly hydro) we checked if THET – the Total number of Hours in which prices were determined by Expensive Technologies, that is Coal, TNC and CC – were less than THCT – the Total number of Hours characterized by Cheap Technologies, that is Hydro and Other – determining instead the daily price. We then computed a kind of “error rate” defined as THET/24 and computed the frequency of days in which this threshold was overcome. According to this test, the previous cases for Hydro (Other in brackets) reduced to 210 (7) in North, 126 (10) in CNorth, 109 (12) in CSouth, 112 (12) in South and 99 (14) in Calb over a sample of 1461 days. Hence, a reader who prefers to focus on cheap/expensive technologies should remind that the *error*

f_{rjt} the number of hour for the r -th technology group used in zone j on day t , the dummy variable for the r -th group in zone j is then defined as

$$\begin{aligned} d_{rjt} &= 1 \text{ if } f_{rjt} = \max_r(f_{rjt}) \\ d_{rjt} &= 0 \text{ otherwise.} \end{aligned} \quad (1)$$

Table 1 gives indications on the frequencies of technologies and justify our decision to exclude two technologies, *Wind* and *Other* in all zones, from our analysis since they had a low influence compared to the other sources.

	Coal	CC	Thermal	Wind	Hydro	Other	Totals
North	73	632	366	0	449	40	1560
CNorth	122	462	702	0	218	26	1530
CSouth	143	362	817	0	183	26	1531
South	151	356	815	0	185	25	1532
Calabria	188	351	810	0	156	28	1533

Table 1: Frequencies of technologies fixing the price over individual zones, considering a sample of 1461 days from 1 January 2005 to 31 December 2008. Different total amounts account for technologies appearing with the same maximum frequency in one day when determining the zonal price

3.2. Market Concentration

The number of operators has increased progressively through years growing from 66 and 76 of the sale side bidding and of the demand side bidding respectively on May 2005 to 98 and 95 sales and purchases operators on December 2008. These numbers refer to participants of both exchange and bilateral markets (24). Participants registered on the IPEX market increased from 51 in 2005 to 151 in 2008 (25). It could be possible to consider the number of market participants because as this number increases the market becomes more competitive and the price should decrease. Hence we expect to observe a reduction of national and zonal prices with the progressive increasing of competition. However this information can be used only at a

rates for hydro (other) are equal to 14.37% (0.48%) in North, 8.62% (0.68%) in CNorth, 7.46% (0.82%) in CSouth, 7.67% (0.82%) in South and 6.78% (0.96%) in Calb. However, we would like to emphasize that even if we did this underestimation in favor of hydro, this technology was not found to be significant in affecting the zonal prices. See section 4.2.1.

national level. Instead we have decided to use the *Residual Supply* index (RSI), computed by GME and available for all market zones¹³.

3.2.1. The Residual Supply Index (RSI)

The Residual Supply Index measures the presence of residual market participants necessary to cover the total demand, thus the index measures the ex-post residuality. The hourly zonal RSI published by GME is computed according to the following formulation

$$RSI_i(j, h) = \sum_{l=1, l \neq i}^N S_l(j, h) - V_i(j, h) \quad (2)$$

where $l, i = 1, \dots, N$ are market participants, j represents the individual zones, h is the considered hour and finally V_i are volumes sold by the i -th participant. This difference between the total supply and the sum of i th sellers' supply (or in other words the quantity offered by other market participants) represents the *non-contestable volumes*. Hence, we have divided these hourly quantities by the total quantities of electricity purchased in one zone at one particular hour, and so we determined the hourly time series of the *true residual supply index*, $TRSI_i$. Later this has been aggregated on daily basis as simple average across 24 hours. Therefore it accounts for both the market power exercised by generators and for electricity demanded on individual zones¹⁴. If the index is less than 1, then the i th firm is necessary to cover the demand and so it is a *pivotal* supplier in the market; if the index is greater or equal to 1, then the i th firm is not necessary and the market can be considered competitive, see Manuhutu and Owen [42] and Rahimi and

¹³It is well-known that the HHI is a traditional structural index which measures static concentration and it represents just one of major sign of market power (see 32). However, when looking at the HHI zonal time series for hours collected in peak and off-peak periods, a sensible shift in level was detected for the entire month of November 2008. This phenomenon was observed neither in the quantities sold (demands) nor in the RSI, for details see 25 page 96. Therefore, we have decided to use only the residual supply index for testing the effect of market power on zonal prices.

¹⁴The formulation of RSI does depend on demand levels and indeed the correlation between the two (daily) series is equal to 0.73. When we divide it by volumes, obtaining the TRSI, there is still a positive correlation but reduced to 0.53. These results refer to *North*. As for Central North, where we have observed $\rho_{RSI, dem} = 0.17$ against $\rho_{TRSI, dem} = -0.04$, finally similar situations for other zones.

Sheffrin [49]. According to the index dynamics, only two are reported for lack of space, we can conclude that all zonal markets cannot be considered competitive¹⁵.

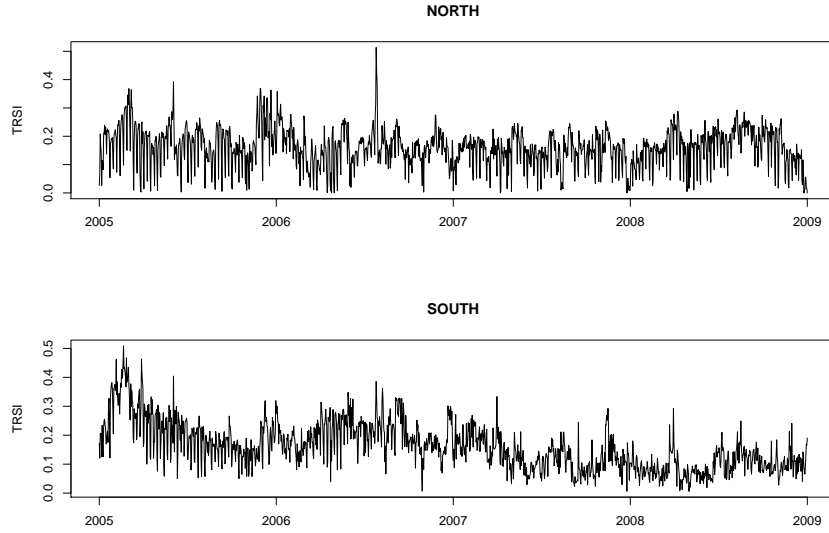


Figure 2: Dynamics of True Residual Supply Indexes for two selected zones (North and South)

¹⁵And this is confirmed by the percentages of HHI levels: 98% of the times the index was above the threshold of 1800 which identifies a situation of *market concentration* for all zones apart *North*, which instead was found to be *moderately concentrated* on 91% of the time (with respect to the sample employed in this study).

3.3. Congestions

A congestion typically implies that a transmission constraint has been hit, so prices on different sides of the constraint may be determined by different technologies and plants. This can be empirically observed by comparing prices of contiguous zones. Hence following [29], we define daily time series of congestions every time that pairs zonal prices are different.

Preliminary investigations performed on couples of zonal daily median prices provided evidence on the importance of congestion state. Following Hadsell and Shawky [28], Gianfreda and Grossi [20] defined the difference between zonal price and the single national price (PUN) as a marginal congestion cost and showed that the Italian market is inefficient since not all zonal prices are equal to the PUN prices. Now, instead of using congestions costs, we identify and define daily time series of frequencies of congestions every time we observe different zonal prices among couples of contiguous zones, for instance we have observed how many times Northern prices were different from Central Northern prices hence indicating a *congestion* between these two zones. All these yearly occurrences are reported in Table 2. In addition, we have added significant frequencies of congestions at all borders adjusting for total hourly congestions¹⁶. We expect then higher prices in zones with high frequencies as consequence of insolation and so unsatisfied demand. For instance, a congestion along the border *North vs. CNorth* implies higher demand, and so prices, in the latter zone; and lower prices in the former one.

4. Model Specifications and Empirical Results

4.1. Preliminary Analysis and Model Selection

A preliminary empirical analysis of the Italian zonal market carried out using daily medians of prices has provided evidence of the presence of seasonality at daily level and a long memory autocorrelation structure.

¹⁶As an example, *CNorth* is connected with *North*, *CSouth* and *Sard*, so in one day we have counted 46 congestions given by adding up frequencies of congestions at all three borders every time that we observed differences between Central Northern prices and those in connected zones. Then we have divided the daily amounts by the daily total possible congestions for that zone, that is by 72 (accounting for 24 hours in a day and for 3 zones). Similarly for the other zones.

Zone	vs	Contiguous zones	2005	2006	2007	2008
North	vs	CNorth	1511	3035	2927	1040
CNorth	vs	North, CSouth and Sard	5296	5639	5552	4687
CSouth	vs	South and CNorth	1048	353	580	1438
South	vs	CSouth and Calb	704	2144	361	587
Calb	vs	South and Sici	5017	6005	4926	6175

Table 2: Total daily frequencies of congestions across interconnected zones observed through studied years

Figure 3 represents correlograms (AutoCorrelation Function, ACF, and Partial AutoCorrelation Function, PACF) for seasonally adjusted prices collected in the Northern zone¹⁷. Seasonal adjustment has been carried out by using a linear model with dummy regressors for days of the week and calendar effects (*CalEf*). The ACF shape is very similar the theoretical pattern of a long memory process, while the first few lags of the PACF are outside the probability bands. This means that the seasonal adjustment did not capture all the weekly seasonal dependence of the series. For this reason we decided to use the original series (two of them are reported in Figure 4 to show similar patterns common to all series considered in this analysis) as dependent variable and explicitly model the seasonality of the series.

In order to test the hypothesis of long memory process, the KPSS test for stationarity and the Philips-Perron test for unit roots have been applied (see Table 3 for results). Both tests¹⁸ reject the null hypothesis confirming the long memory process as a possible generating process of the prices series.

Long memory is usually captured by fractionally integrated processes. Then, taking into account the autocorrelation structure we could estimate ARFIMA models with seven Autoregressive terms, that is a ARFIMA(7,0) or seven Moving Average terms, that is a ARFIMA (0,7). Another stylized fact that should not be neglected is the mean reversion of electricity prices

¹⁷Similar dynamics are observed for all other zones and are not reported here for lack of space.

¹⁸A significant debate has considered the properties of tests for unit root, or long memory in the presence of structural breaks. It has been shown that persistence tests are severely compromised, in terms of their size and power properties, in series which display breaks (1). However, there are no reasons to consider the presence of structural breaks in the Italian market in the considered time period.

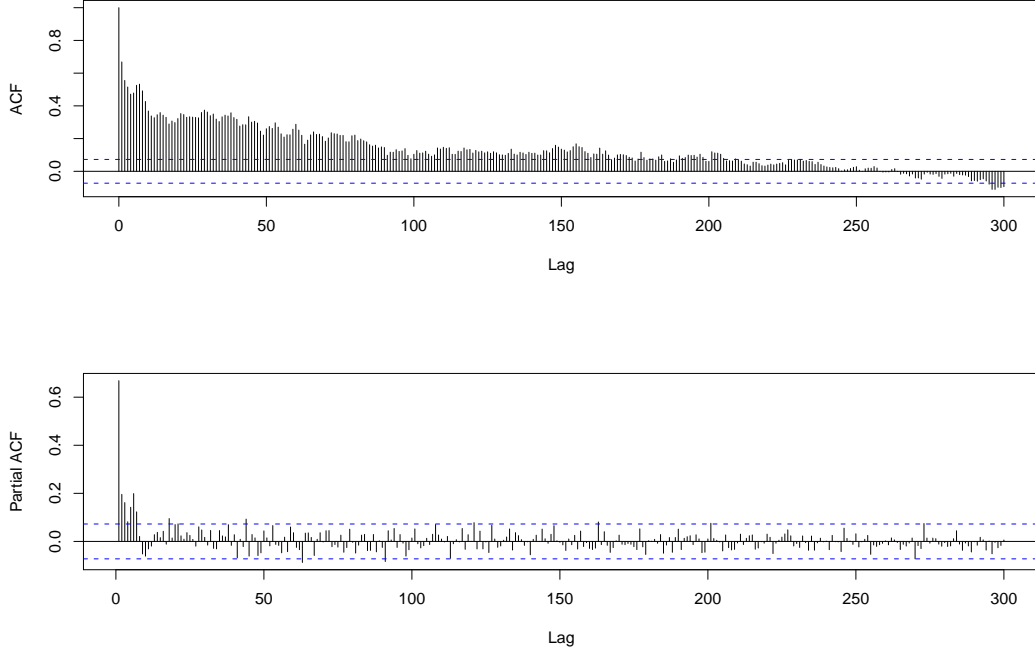


Figure 3: Autocorrelation functions of seasonally adjusted prices for North

	North	CNorth	CSouth	South	Calb
PP test	-19.859	-20.565	-19.086	-19.154	-20.006
KPSS test	0.435	0.295	0.406	0.395	0.389

Table 3: Stationarity (KPSS) and unit roots (PP) tests for daily prices. Thresholds at 1% and 5% level are -3.970 and -3.415 for PP test; and 0.216 and 0.146 for KPSS test

which is usually estimated by a one-lag autoregressive term. For this reason we added one AR term also in the case of the ARFIMA model with seven Moving Average terms. Finally, the two estimated ARFIMA models are: ARFIMA(7,0) and ARFIMA(1,7). Residual diagnostics¹⁹ from these models show that ACF and PACF functions are inside the confidence regions, but

¹⁹Tables are not reported but are available on request.

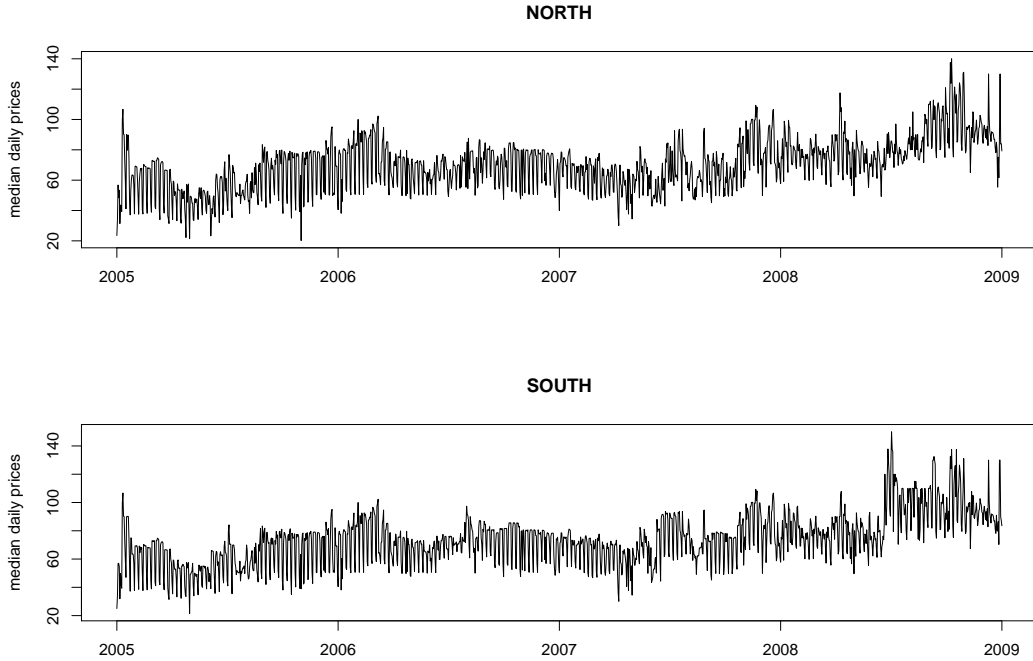


Figure 4: Dynamics of Daily Median Prices for North and South.

the null hypothesis of homoscedasticity, according to the Engle LM test, cannot be accepted. For this reason, we estimated ARFIMA–GARCH models and residuals diagnostics, reported in Table 4, lead to the acceptance of the ARFIMA(1,7) combined with a GARCH(1,1) against an EGARCH(1,1) model, hence supporting the evidence against a *leverage effect* characterizing Italian prices, provided in [18].

The best model, the ARFIMA(1,7)–GARCH(1,1) with a Student–t distribution²⁰ for residuals, selected according to the information criteria (AIC and BIC), has been used as the basic model for testing the influence of ex-

²⁰We estimated the models under the assumption of different distributions for residuals to take into account the presence of many extreme values and consequent fat tails of the distribution of electricity prices. The best performance has been obtained using a Student–t distribution.

	NORTH	CNORTH	CSOUTH	SOUTH	CALB
Q-Statistics on Standardized Residuals					
Q(15)	0.302	0.406	0.340	0.369	0.251
Q(20)	0.234	0.560	0.656	0.724	0.454
Q(30)	0.112	0.373	0.242	0.394	0.165
Q-Statistics on Squared Standardized Residuals					
Q(15)	0.581	0.915	0.704	0.627	0.749
Q(20)	0.640	0.972	0.718	0.609	0.775
Q(30)	0.635	0.552	0.042	0.715	0.929
Diagnostic test based on the news impact curve (EGARCH vs. GARCH)					
Sign Bias Test	0.542	0.594	0.459	0.274	0.535
Negative Size Bias Test	0.851	0.620	0.820	0.674	0.835
Positive Size Bias Test	0.712	0.780	0.719	0.682	0.512
Joint Test	0.749	0.891	0.720	0.479	0.618
LM Engle test					
ARCH 1–2 test	0.742	0.984	0.851	0.750	0.744
ARCH 1–5 test	0.390	0.990	0.362	0.294	0.342
ARCH 1–10 test	0.575	0.835	0.475	0.385	0.511

Table 4: P-values of residuals diagnostic tests for ARFIMA(1,7)–GARCH(1,1) models estimated in five zones

planatory variables on wholesale zonal prices. Hence, the effect of exogenous factors on wholesale prices has been measured implementing Reg-ARFIMA-GARCH models, as in [39], with dummies for groups of technologies and for congestions, the computed index for market concentration and traded volumes. In the next section, the model specification will be formalized, while exploring the relation between prices and individual explanatory variables within the framework of the ARFIMA-GARCH processes. The combined effects²¹ of most significant variables will be then estimated in subsection 4.2.5.

4.2. Model Specifications

The proposed models can be formalized as follows:

$$\Phi(L)(1-L)^d(y_t - \mu_t) = \Theta(L)\varepsilon_t \quad \varepsilon_t|I_{t-1} \sim t(0, \sigma_t^2) \quad (3)$$

with

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \quad (4)$$

for $t = 1, \dots, T$, y_t is the zonal median electricity price at time t , L is the lag operator defined by $Ly_t = y_{t-1}$ and $\mu_t = E(y_t|I_{t-1})$ is the mean equation conditioned to the set of information available at time $t - 1$.

The operators $\Phi(L) = 1 - \sum_{i=1}^p \phi_i L^i$ and $\Theta(L) = 1 + \sum_{j=1}^q \theta_j L^j$ are the autoregressive and moving average polynomials, respectively. Accordingly to previous discussion, the best model for Italian prices is an ARFIMA(1,7), so that $p = 1$ and $q = 7$.

The following specification has been considered for the conditional mean function:

$$\mu_t = \mu + \nu_1 D_t^1 + \dots + \nu_6 D_t^6 + \gamma CalEf_t + \lambda'_v x_t \quad (5)$$

where D_t^j with $j = 1, \dots, 6$ are dummies for days of the week and ν_j are the corresponding coefficients; $CalEf_t$ is a dummy accounting for calendar effects and γ is the corresponding coefficient; x_t represent a vector of explanatory variables with $v = 1, 2, 3, 4$ indicating respectively technologies determining the price when $v = 1$ and $x_t = Tech_t$; the indexes of market power when $v = 2$ and $x_t = MarPow_t$; congestions for daily events when $v = 3$ and $x_t = Cong_t$; and finally electricity volumes when $v = 4$ and

²¹The analysis of the correlations between explanatory variables produced no significant results. Tables are available from the authors on request.

$x_t = Vol_t$. λ s are vectors of regression coefficients. The size of x_t and λ varies with v . Therefore we have initially tested effects of single groups of explanatory variables to understand their implications on zonal prices, and then we have selected the significant ones to verify their combined effects on prices. All Reg-ARFIMA(1,7)-GARCH(1,1) parameters have been obtained by one-step maximum likelihood estimators using the statistical language *Oct* and the internal functions of G@RCH (version 5.0).

4.2.1. Effects of Technologies

	Technology	Coef	Std. Err.	t-stat	p-value	AIC	
Calb	Coal	-1.073	0.590	-1.819	0.069	6.952	(*)
	CC	-2.507	0.655	-3.827	0.000	6.926	(***)
	TNC	3.031	0.723	4.194	0.000	6.919	(***)
	Hydro	0.514	0.652	0.788	0.431	6.955	
	Other	1.466	1.849	0.793	0.428	6.954	
CNorth	Coal	-1.110	0.635	-1.747	0.081	6.845	(*)
	CC	-1.177	0.522	-2.256	0.024	6.840	(**)
	TNC	2.113	0.698	3.026	0.003	6.831	(***)
	Hydro	0.048	0.606	0.080	0.937	6.849	
	Other	1.597	2.736	0.584	0.560	6.847	
CSouth	Coal	-0.737	0.598	-1.233	0.218	6.927	
	CC	-2.272	0.661	-3.436	0.001	6.903	(***)
	TNC	2.852	0.725	3.935	0.000	6.894	(***)
	Hydro	0.040	0.628	0.064	0.949	6.928	
	Other	0.977	1.680	0.582	0.561	6.928	
North	Coal	-1.334	1.001	-1.332	0.183	6.902	
	CC	-2.491	0.606	-4.108	0.000	6.875	(***)
	TNC	3.849	0.999	3.853	0.000	6.873	(***)
	Hydro	0.791	0.649	1.219	0.223	6.902	
	Other	2.347	1.609	1.458	0.145	6.900	
South	Coal	-0.503	0.581	-0.866	0.387	6.931	
	CC	-2.321	0.659	-3.523	0.001	6.906	(***)
	TNC	2.877	0.733	3.925	0.000	6.899	(***)
	Hydro	0.035	0.649	0.054	0.957	6.932	
	Other	1.567	1.903	0.823	0.411	6.930	

Table 5: Estimates of Reg-ARFIMA-GARCH models for the only effects of technologies

To control for technologies, we have estimated a Reg-ARFIMA(1,7)-GARCH(1,1) with $x_t = Tech_t$ in the conditional mean equation, where $Tech_t$ is a vector of zeros and ones obtained according to (1). Looking at Table 5, it is possible to draw the following comments on the employed *groups of technologies* determining zonal prices. First of all, *Hydro* and *Other* are never found to be significant in all zones, so even if determining the zonal prices these factors are uninfluential to the price dynamics. *Coal* is found to be marginally significant (at 10% level) only in CNorth and Calabria, reducing the corresponding prices; on the contrary *Combined Cycles*, (CC), and *Thermal* power, (TNC), are always significant. In details, the former technology reduces electricity zonal prices, whereas the latter ones increase them. According to these results, we have selected only the last two variables to be included in the final model formulation.

4.2.2. Effects of Market Concentration

Concentration has been investigated using the *true residual supply* index in the conditional mean, that is $x_t = TRSI_t$ in eq. (5). Results reported in Table 6 confirm our expectations: the lack of competition in all zones affects prices inducing increased values. Then it is important to include this variable in our final model to account for the exercise of market power.

	Coef	Std. Err.	t-stat	p-value	AIC	
Calb	0.204	0.105	1.947	0.052	6.952	(*)
CNorth	0.123	0.040	3.081	0.002	6.832	(***)
CSouth	0.046	0.024	1.901	0.058	6.921	(*)
North	0.837	0.087	9.638	0.000	6.726	(***)
South	0.254	0.074	3.411	0.001	6.910	(***)

Table 6: Estimates of the only effects of market power, measured by TRSI in the Reg-ARFIMA-GARCH models

4.2.3. Effects of Congestions

To control for congestions, we have estimated the selected model with $x_t = Cong_t$ in the conditional mean equation, where $Cong_t$ is a vector of frequencies of *congestions*. Considering Table 7, we find that *congestions* along borders do affect zonal prices but sometimes with surprising signs. Accounting for the direction of electricity flows, we expect that zonal prices will

be affected by congestions when electricity flows versus the considered zone. Therefore, first of all we observe that Northern prices are affected by congestions, but interestingly these decrease zonal prices (at 1% confidence level) and it can be argued that, when congestions occur, the Northern demand could be satisfied by imports from foreign markets. Conversely, we find that congestions do not affect Central Northern prices, when electricity flows in the opposite direction from North to Central North. And similarly for congestions caused by flows from Central South, whereas those congestions on flows arriving from Sardinia do increase zonal prices in CNorth. Congestions increase CSouthern prices, as well as prices in Calabria when the lines with South are congested. But Calabrian prices decrease when there is a congestion with Sicily since here another limited production pole is based, Rossano. Finally, Southern prices are not affected by congestions and the motivation could lie in the presence of a limited production pole, Brindisi, which injects electricity into the system even if the zone is congested and so separated from Calabria and CSouth. Only significant congestion events will be considered in the following analysis.

Zones	Flow Direction	Coef	Std. Err.	t-stat	p-value	AIC
North	North←CNorth	-3.904	0.736	-5.303	0.000	6.858 (***)
CNorth	North→CNorth	0.827	0.636	1.300	0.194	6.846
	CNorth←CSouth	-1.601	1.117	-1.433	0.152	6.845
	CNorth←Sard	1.053	0.560	1.881	0.060	6.844 (*)
CSouth	CNorth→CSouth	3.333	1.706	1.954	0.051	6.917 (*)
	CSouth←South	2.261	1.234	1.832	0.067	6.927 (*)
South	South←Calb	0.662	1.368	0.484	0.628	6.932
	CSouth→South	1.184	2.646	0.448	0.655	6.932
Calb	South→Calb	3.310	1.787	1.852	0.065	6.949 (*)
	Calb←Sici	-0.911	0.469	-1.941	0.053	6.951 (*)

Table 7: Estimates of the only effects of congestions on zonal prices in the Reg-ARFIMA-GARCH models

4.2.4. Effects of Volumes

Volumes are considered as the best and the most used explanatory variable. Thus, in line with the literature, we have also tested empirical relations

between zonal prices and volumes. Looking at Table 8, it is possible to observe that in all zones estimated coefficients are significant and positive, hence when demand rises, also prices increase as a consequence of calling for additional generation from more expensive units. Therefore, this explanatory variable will be included into the final formulation²². It is worth noticing that the highest coefficient is observed in Calabria where electricity is mainly produced by thermal units, while the lowest coefficient is observed in the North where the majority of hydro units is.

	Coef	Std.Err	t-stat	p-value	AIC	
North	0.098	0.013	7.420	0.000	6.829	(***)
CNorth	0.665	0.077	8.690	0.000	6.728	(***)
CSouth	0.871	0.113	7.710	0.000	6.834	(***)
South	0.423	0.068	6.201	0.000	6.863	(***)
Calb	2.153	1.040	2.069	0.039	6.947	(**)

Table 8: Estimates of the effects of observed volumes on zonal prices in the Reg-ARFIMA-GARCH models

4.2.5. Combined Effects and Preliminary Comments

After the appropriate variable selection, we have tested the model (3)-(5) for all significant explanatory variables, that is considering the following formulation for the conditional mean:

$$\mu_t = \mu + \nu_1 D_t^1 + \dots + \nu_6 D_t^6 + \gamma CalEf_t + \lambda'_1 Tech_t + \lambda'_2 MarPow_t + \lambda'_3 Cong_t + \lambda'_4 Vol_t. \quad (6)$$

According to estimated parameters, reported in Table 9, we can draw the following preliminary conclusions:

1. *calendar effects, seasonality, fractional integration* as well as *volatility clustering* are important and salient features to take into account since the estimates – *CalEf*, days of the week, d , α and β – are always significant. Moreover d is less than 0.5 for all zones, as found previously in [20], hence confirming that these price processes have long memory;

²²The correlations observed between volumes and other variables are low, and in details that one with TRSI, equal to 0.53, does not indicate problems of multicollinearity.

2. the *autoregressive structure*, that is the ϕ_1 term, is not found to be significant. One possible explanation could be that volumes capture effects which are usually explained by the past history of the price series. The inclusion of generally significant moving average terms has been used to obtain white noise residuals;
3. the employed *groups of technologies* determining the zonal prices are generally significant across zones. And the final formulation confirms that *Combined Cycles* (CC) always reduce electricity zonal prices, whereas *Thermal power* (TNC) generally increases them;
4. *concentration*, analyzed using the TRSI, shows a positive sign in all considered zones implying that the residual supply was not sufficient to cover zonal demand hence inducing prices to increase. Interestingly, in the final formulation it turned out to become non-significant in South;
5. *congestions* are important only in North and CSouth, but with different signs. In the first case, it can be argued that when congestions affect the Northern zone, demand could be satisfied by imports. On the other hand, when congestions occur in CSouth electricity prices raise because of an excess of demand and problems of market concentration. Now prices in CNorth but especially in Calabria turn not to be influenced by congestions;
6. finally, *volumes* are confirmed to be always significant and positive in all zones.

	CALB		CNORTH		CSOUTH		NORTH		SOUTH	
μ	15.624		6.073		-7.166		36.328	(***)	19.286	(***)
TRSI	0.189	(*)	0.089	(**)	0.042	(**)	0.655	(***)	0.092	
VOL	2.143	(**)	0.632	(***)	0.846	(***)	0.055	(***)	0.366	(***)
Cong	2.033		0.808		3.028	(***)	-3.037	(***)	0.742	
CC	-1.567	(**)	-0.324		-1.221	(**)	-1.354	(**)	-1.480	(**)
TNC	2.267	(***)	1.102		1.585	(**)	1.325		1.459	(**)
CalEf	-19.577	(***)	-7.548	(***)	-8.036	(***)	-6.913	(***)	-16.401	(***)
Mon	16.613	(***)	3.949	(*)	4.512	(**)	1.803		12.838	(***)
Tue	17.635	(***)	2.489		2.936		3.110		12.549	(***)
Wed	17.651	(***)	2.019		2.754		2.640		12.569	(***)
Thu	16.979	(***)	1.113		1.877		2.032		11.568	(***)
Fri	16.011	(***)	1.061		1.085		1.120		10.459	(***)
Sat	4.371	(***)	-2.282	(*)	-2.819	(**)	-0.685		1.681	
d	0.480	(***)	0.464	(***)	0.488	(***)	0.473	(***)	0.491	(***)
ϕ_1	0.191		0.064		0.105		0.138		0.089	
θ_1	-0.162		-0.178		-0.109		-0.268	(**)	-0.113	
θ_2	-0.113	(**)	-0.110	(**)	-0.140	(***)	-0.023		-0.141	(***)
θ_3	-0.074	(*)	-0.047		-0.053		-0.027		-0.079	(*)
θ_4	0.020		-0.009		-0.008		-0.013		-0.015	
θ_5	-0.012		-0.008		0.006		-0.038		-0.003	
θ_6	0.015		0.097	(**)	0.043		0.041		0.027	
θ_7	0.222	(***)	0.182	(***)	0.232	(***)	0.182	(***)	0.232	(***)
ω	1.867		1.919	(*)	1.452		0.995		2.193	(*)
α	0.091	(***)	0.105	(***)	0.088	(***)	0.059	(***)	0.113	(***)
β	0.888	(***)	0.871	(***)	0.896	(***)	0.924	(***)	0.860	(***)
t-Stud	5.360	(***)	4.687	(***)	4.920		6.691	(***)	5.602	(***)

Table 9: Reg-ARFIMA estimates (with p-values in brackets) for Italian Electricity Zonal Prices

5. Assessing forecasting performance

In this section the forecasting performance of Reg-ARFIMA-GARCH model with conditional mean equation specified in (5) is investigated. Given that dependent variable and regressors are inserted into the model at the same time, we cannot obtain out-of-sample predictions because at time t the values of regressors are unknown. When explanatory variables are binary (as technologies and congestions), a scenario analysis can be applied and so described in section 5.1; whereas when variables are continuous (as TRSI and volumes), predicted values are necessary. The Italian independent system operator, GME, provides day-ahead predictions for demand. Instead, predicted values for TRSI are not available, so we decided to exclude this variable in the prediction of prices leaving this issue as future development. To summarize, we modify model (5) by removing the TRSI and replacing observed volumes with the forecasted demand to evaluate the forecasting performance. This model will be called “Final Model”. We assume to have knowledge of history up to the end of June 2008 and try to assess the performance ability of the model. In other words, we use daily data from 01/01/2007 until 30/06/2008 as a sort of “training data set” and measure the forecasting performance of the model until the end of 2008.

To evaluate the out-of-sample forecasting performance of the models we use a “rolling windows” procedure. In order to make clear how this technique works, the whole time period is divided in two sub-periods, the first going from $t = 1$ to $t = T - m$ and the second covering the period from $t = T - m + 1$ to T . The procedure is iterative as we use a different set of information for estimating purposes rolling a windows of $T - m$ observations over the original data-set. Every time the estimated parameters are used to get a one-step-ahead forecast. Going into details, the rolling windows procedure works as follows:

- at time $T - m$ the vector of estimates θ_{T-m} is obtained through different models (RW1, RW7, Basic and Final) using data for $t = 1, \dots, T - m$; the h -ahead forecast in $T - m + 1$, is then given by

$$y_{T-m+h|T-m} = f(\theta_{T-m}, y_{T-m}).$$

- at time $T - m + 1$ the forecast for time in $T - m + h + 1$ is obtained on data for $t = 2, \dots, T - m + 1$, that is

$$y_{T-m+h+1|T-m+1} = f(\theta_{T-m+1}, y_{T-m+1})$$

- ...
- the last forecast is estimated at time $T-h$, using data for $t = d, \dots, T-h$

$$y_{T|T-h} = f(\theta_{T-h}, y_{T-h}).$$

where $d = m - h$. At the end of the iterative procedure, $m - h + 1$ h -step ahead forecasts are obtained. Analyzing daily data of electricity prices, if $m = 180$, we can check the h -day ahead forecasting performance for the last six months of data.

To evaluate the gain obtained by using exogenous variables (the “Final Model”) we considered three benchmark models: a simple random walk (RW1), a weekly random walk (RW7) and the ARFIMA(1,7)–GARCH(1,1) model without regressors (from now on, called “Basic Model”). The RW1 is a classical benchmark model whose forecasts are commonly called “naive” predictions. The forecast function of the random walk is $y_{t|t-1} = y_{t-1}$, that is the observed average price of yesterday is the forecast for today. We take the value of two days ago if there was a holiday yesterday. The number of days in the past is increased accordingly when there are two or more contiguous holiday days. The RW7 is a forecast method which has been used as benchmark model in previous papers on electricity loads forecasting (as in 52; 12). The forecast function for the RW7 is $y_{t|t-1} = y_{t-7}$, that is the average price observed one week ago is the forecast for today. We take the value of two weeks ago if there was a holiday one week ago. Anyway special days, holidays included, arise many problems, thus we have deleted these forecasts for the RW1 and RW7 since we consider them only as benchmark models. The *Basic Model* is a restricted version of the *Final Model*. The comparison between the Basic and the Final models is carried out to evaluate how the exogenous variables can improve the forecasting performance. This is a crucial point because, comparing the two models, we can properly evaluate whether the complexity of the Final model is justified. If the Final model will not perform significantly better than the Basic, then the simple knowledge of past prices is enough for predictive purposes and exogenous variables are completely useless.

Figure 5 presents the one-day-ahead relative forecast errors (forecast errors divided by observed prices) for the North macro-region as well as their

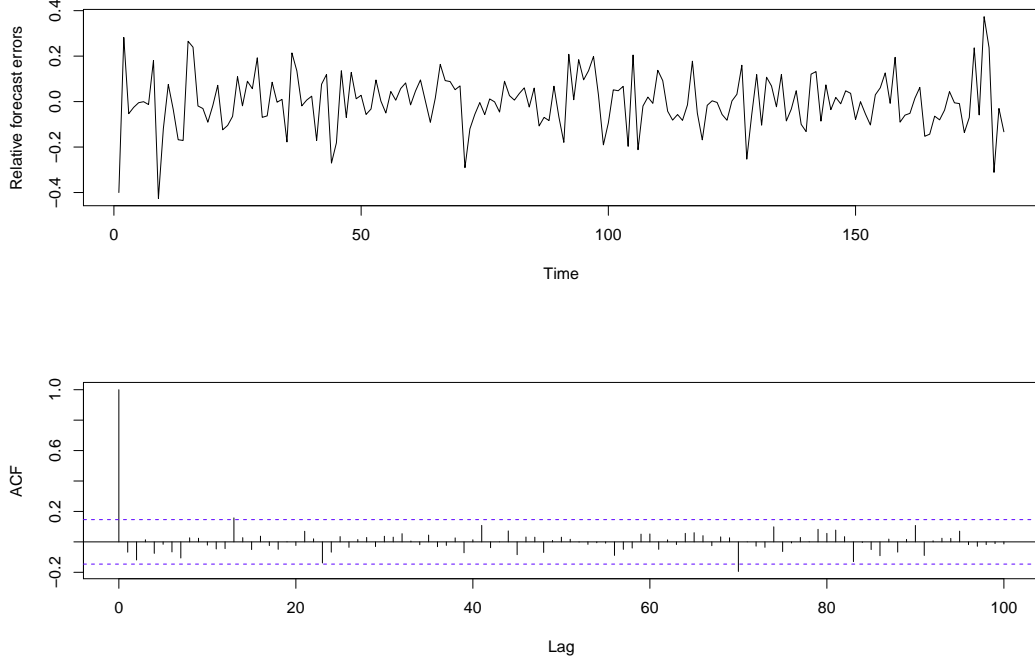


Figure 5: Out-of-sample one-step-ahead relative forecast error (upper panel) and corresponding empirical autocorrelation function (lower panel) for North region according to the *Final Model*

autocorrelation function using the Final model²³. On the whole, the forecasts seem unbiased. The largest relative errors correspond to the final days of 2008. The correlations are slightly outside the confidence region only for few lags, but in general the ACF plot does not show any particular pattern that could be considered as a clue of a missed structural dynamic feature in the time series.

To assess the predictive goodness of each model, we use a set of measures: the root mean squared forecast error (RMSE), the mean absolute percentage forecast error (MAPE), and the Theil's U index. We also apply the Diebold–Mariano test (Diebold and Mariano, 1995) to compare different estimated models.

²³Plots for the remaining areas are very similar and are available upon request.

		1 step-ahead			
		RW1	RW7	Basic	Final
NORTH	RMSE	14.072	13.001	10.710	9.644
	MAPE	10.696	10.103	8.384	7.366
	Theil's U	-	0.892	0.759	0.685
	DM	-	0.693	2.393	2.022
	p-value	-	0.244	0.008	0.022
CNORTH	RMSE	15.852	14.789	11.563	10.606
	MAPE	11.603	11.089	9.192	8.284
	Theil's U	-	0.965	0.751	0.669
	DM	-	0.571	2.344	2.235
	p-value	-	0.284	0.010	0.013
CSOUTH	RMSE	17.567	17.506	13.310	12.143
	MAPE	12.634	13.402	9.937	9.163
	Theil's U	-	0.942	0.747	0.691
	DM	-	1.020	3.090	3.016
	p-value	-	0.154	0.001	0.001
CALB	RMSE	16.947	17.233	12.886	12.336
	MAPE	12.382	13.319	9.689	9.067
	Theil's U	-	0.969	0.758	0.723
	DM	-	0.853	3.339	3.031
	p-value	-	0.197	0.000	0.001
SOUTH	RMSE	17.067	17.585	13.172	11.959
	MAPE	12.541	13.459	9.829	8.984
	Theil's U	-	0.975	0.765	0.706
	DM	-	0.530	3.327	2.653
	p-value	-	0.298	0.000	0.004

Table 10: Assessment for one-day-ahead forecasts. The following indexes are reported: Root Mean Squared forecast Error (RMSE), Mean Absolute Percentage forecast Error (MAPE), Theil's inequality coefficient (Theil's U), Diebold and Mariano test (DM) and corresponding p-value. As we use the one-sided DM test to evaluate the superiority of one model, the null hypothesis of equal performance is rejected at 1% level when $|DM| > 2.33$.

Table 10 shows the forecast assessment indexes for one-day-ahead forecasts obtained from the four models cited above and for each macro-region. The MAPE and RMSE of the Basic and Final models are always far lower than those of the Random Walk models. Moreover, the Theil's index always shows that considering the autocorrelation structure of the time series always lead to an improvement of the forecasts with respect to the "naive" model. Finally, the last two rows for each zone report the DM tests and the corresponding p-values. Models are compared by couples moving from the simplest to the more complex model. In details, the RW7 model is compared with the RW1 model in the second column; whereas in the third column, the Basic model is compared with the RW7; and in the last column, the Final model is compared with the Basic one. The results of the DM test can be summarized as follows: the RW7 model is never better than the RW1; at 1% level of significance, the Basic models is always better than the RW7; and the Final model is better than the Basic in all zones but North and CNorth where the p-values are slightly greater than 0.01 (0.022 and 0.013, respectively). This weak difference in the forecasting performance between Central-Northern and Southern Italy could be due to the fact that congestions, combined cycles and thermal units turn out to be not significant and so inducing no contribution to the forecasting ability of the overall model in CNorth; whereas this could be due simply to thermal units in North.

At this point we must remember that, since the Final model includes explanatory variables, the forecasting accuracy could be based either on realized daily values of the explanatory variables or on their one-day-ahead forecasts. There are pros and cons in both approaches. The former (using realized values) may be preferred to avoid having to discuss external inaccuracies due to explanatory variables forecast errors, while the latter (using forecasted values of explanatory variables) may be preferable since the corresponding model could be used as in real situations. Anyway, a scenario analysis could be applied in case of binary variables such as congestion events and technologies. For instance, we could compare the models predicting the price for tomorrow assuming a congestion event and a given production technology. The scenario approach is discussed in section 5.1.

5.1. Scenario analysis

In this section a scenario analysis for price forecasting is introduced. Taking into account that technologies and congestions are binary variables, we can simulate for each future day a upper and a lower boundary of predicted

prices considering the sign of the corresponding coefficients. Thus, we could predict the price for tomorrow assuming a congestion event and a given production technology, other things being equal. For instance, in the Northern zone, the sign of coefficients are negative for CC and Congestions, while TNC has a positive sign. This means that, *ceteris paribus*, the upper bound of the forecasted price will be obtained when the dummies for CC and Congestion will be zero and the dummy for TNC will be equal to one (this can be called as “0–0–1” scenario indicating the worst scenario for higher prices). Conversely, the lower boundary is obtained for CC and Congestion equal to one and TNC equal to zero (this can be called as “1–1–0” scenario indicating the best scenario for lower prices). Results reported in Figure 6 for the Northern zone are an example of how the scenario procedure work. The dashed line is for one-day-ahead forecast in the first sixty days after $T - m$, obtained through the rolling window procedure²⁴. The continuous lines are obtained by means of the scenario analysis, replacing the observed values of congestions and technologies with the 0–0–1 scenario (upper bound) and the 1–1–0 scenario (lower bound). Therefore, an out-of-sample predicted interval for price in $t + 1$ is obtained at each time t . As it can be noticed, predictions according to our Final Model are always in the predicted interval and so providing evidence that our forecasts are actually between maximum and minimum possible prices. Similar results are found for other zones and available on request.

6. Conclusions

The understanding from our analysis is important for drawing policy indications because, instead of considering only volume as done in previous studies, other three variables – technologies, congestions and concentration – immediately provide some specific policy suggestions. In details, they could be summarized as follows:

1. *Proper price dynamics modeling*: the identification of the best process generating electricity prices (see points 1 and 2 in Section 4.1) will help in understanding the deregulation process, verifying the degree of competition as well as the foreseeability of such prices. This provides

²⁴In this case, the one-day-ahead value of forecasted demand is known, while congestion and technologies values are not, thus observed values are used.

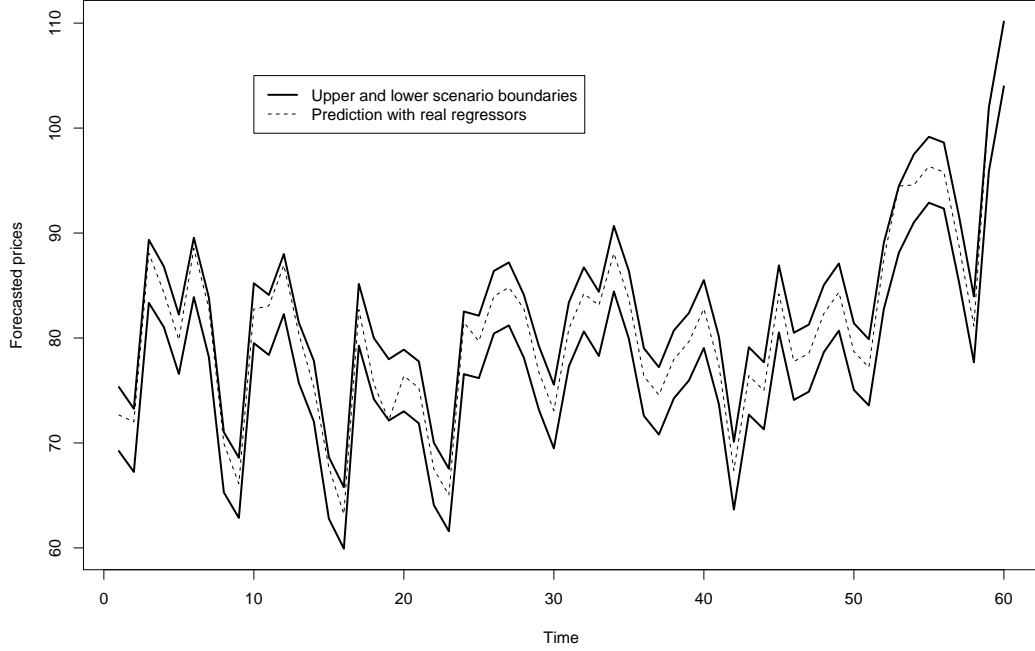


Figure 6: Predicted prices in the North Zone with upper and lower boundaries superimposed

specific information to market participants on how to evaluate their optimal bidding strategies on a short-term period, or how to base bilateral contracts on a longer term, or finally on spot and forward price definitions and relationships with serious implications for derivative pricing and risk-hedging problems.

2. *Congestions management*: a good model identification leads to proper managing of intra-network congestions for needs of continuous real time balancing. Moreover, *new investments* in constructing new grid interconnections or reinforcing the existing ones will increase the transmission capacity within the market (that is across zones) or between markets (that is across and with foreign markets) leading to reduced wholesale prices, a more integrated area with less market price manipulations, and consequently improving market competitiveness and efficiency.

3. *Production planning*: great attention must be given to generation sources, the optimal technology mix, the construction of new power plants together with the monitoring of plants availability. It is worth mentioning that Italy does not have nuclear power generation and invested (until the end of 2008) in wind power as France did, but less than Germany: 3736 MW against 3404 MW in France, where the main generation source is nuclear power; and against 23903 MW in Germany and 16740 MW in Spain, see [14] for more insights. In addition, in those years EU experienced an increase in natural gas, hydro, biomass and wind power, whereas the last two sources were unable to determine the Italian prices. Therefore more investments are required in alternative resources also in view of the 20–20–20 EU target for carbon emissions. To be precise, Italy invested in 120.2 MW on photovoltaic power in 2007, whereas again Germany and Spain invested respectively 3862 MW and 655 MW, see [34] for additional details. And recently, some interest has been attracted by electricity storage, via pumping stations, as complement to intermittent generation (see for instance 9) and as a possible alternative to costly grid expansion plans. However given that in Italy it is not always possible to combine hydro and wind generation, and so store electricity, regulators should try at least to reinforce network capability.
4. *Market monitoring of the exercise of market power*: regulatory authorities should indeed directly act and drive the construction and the computation of suitable indexes able to account for all, or at least *most*, special market features and therefore working to define new indexes better than those commonly used and implemented.

Therefore, we have provided some indications based on empirical analysis, accounting for the intrinsic market structure with its own generating park and network grid. Indeed this paper quantifies the effects of technologies, concentration, congestions and demand on Italian Electricity zonal prices.

According to the most recent contributions in time series analysis applied to electricity prices, we took into account the long memory feature of the generating stochastic process estimating a parameter of fractional integration, which turned out to lie very close to 0.5. A causal analysis in the framework of Reg–ARFIMA–GARCH models confirmed the significant impact of production technologies, market concentration and congestions on these price dynamics. Moreover, we tested the forecasting ability of sev-

eral models, hence showing that the performance is improved including the selected explanatory variables.

Concluding, we have provided firstly insights on relationships between zonal electricity spot prices, technologies, concentration, congestions and volumes; and secondly, proved that Reg-ARFIMA-GARCH models with exogenous variables perform better than other models in forecasting zonal prices. These results can have important implications when programming the medium-long term energy policy in Italy or future investment strategies with respect to the technology mix and, especially, the network grid since generators can serve only if an adequate transmission capacity exists. Otherwise, the installation of new power plants is expected to produce even more and sudden bottleneck problems.

Acknowledgements

The authors would like to thank the Market Monitoring Area at the GME and in particular the director Cosimo Campidoglio, Pia Saraceno, Sandro Sapia and participants at the *IEFE-FEEM* seminar at Bocconi University and at EUI in Fiesole. Special thanks go to Derek W. Bunn, Mike Clements and two anonymous referees for useful comments which lead to improve substantially the paper. The first author would like to thank the RSCAS at EUI for financial support during the revision process. The work was also partially financed by the Italian Ministry for Education and University with the Project “PRIN 2008 – Robust methods for the analysis of electricity markets”.

References

- [1] Banerjee A. and Urga G., (2005), *Modelling structural breaks, long memory and stock market volatility: an overview*, Journal of Econometrics, 129, pp. 1–34
- [2] Bask M. and Widerberg A., (2009), *Market structure and the stability and volatility of electricity prices*, Energy Economics, 31, pp. 278–288
- [3] Battaglia F. and Orfei L., (2005), *Outlier Detection and Estimation in NonLinear Time Series*, Journal of Time Series Analysis, 26, 1, pp. 107–121
- [4] Blumsack S., Perekhodtsev D. and Lave L.B., (2002), *Market Power in Deregulated Wholesale Electricity Markets: Issues in Measurement and the Cost of Mitigation*, The Electricity Journal, 15, pp.11–24
- [5] Bosco B., Parisio L. and Pelagatti M., (2007), *Deregulated Wholesale Electricity Prices in Italy: An Empirical Analysis*, International Advances Economic Research, 13, pp. 415–432
- [6] Bowden N. and Payne J.E., (2008), *Short term forecasting of electricity prices for MISO hubs: Evidence from ARIMA–EGARCH models*, Energy Economics, 30, pp. 3186–3197
- [7] Conejo A.J., Contreras J., Espinola R. and Plazas M.A., (2005), *Forecasting electricity prices for a day-ahead pool-based electric energy market*, International Journal of Forecasting, 21, 3, pp. 435–462
- [8] Cuaresma J.C., Hlouskova J., Kossmeier S. and Obersteiner M., (2004), *Forecasting electricity spot-prices using linear univariate time-series models*, Applied Energy, 77, pp. 87–106
- [9] Denault M., Dupuis D. and Couture–Cardinal S., (2009), *Complementarity of hydro and wind power: Improving the risk profile of energy inflows*, Energy Policy, 37, pp. 5376–5384
- [10] Diebold F.X., Mariano R.S. (1995), *Comparing predictive accuracy*. Journal of Business and Economic Statistics, 13, pp. 253–263.

- [11] Diongue A.K., Gugan D. and Vignal B., (2009), *Forecasting electricity spot market prices with a k-factor GIGARCH process*, Applied Energy, 86, pp. 505–510
- [12] V. Dordonnat V., Koopman S.J., Ooms M., Dessertaine A. and Collet J., (2008), *An hourly periodic state space model for modelling French national electricity load*, International Journal of Forecasting, 24, pp. 566–587
- [13] Escribano A., Peña J., Villaplana P., (2002), *Modelling electricity prices: international evidence*, Economic Series 08, Working Paper 02-27, Universidad Carlos III de Madrid
- [14] EWEA, The European Wind Energy Association, (2008), *European Wind Map*, document available at http://www.ewea.org/fileadmin/ewea_documents/documents/statistics/European_Wind_Map_2008.pdf
- [15] Fridolfsson S. and Tangeras T., (2009), *Market power in the Nordic electricity wholesale market: a survey of the empirical evidence*, Energy Policy, 37, pp. 3681–3692
- [16] Furió D. and Lucia J.J., (2009), *Congestion management rules and trading strategies in the Spanish electricity market*, Energy Economics, 31, pp. 48–60
- [17] Garcia A. and Arbelaez L.E., (2002), *Market power analysis for the Colombian electricity market*, Energy Economics, 24, pp. 217–229
- [18] Gianfreda A., (2010), *Volume and Volatility Effects in European Electricity Markets*, Economic Notes, 39, pp. 47–63
- [19] Gianfreda A. and Bunn D., (2010), *Integration and Shock Transmissions across European Electricity Forward Markets*, Energy Economics, 32, 2, pp. 278–291
- [20] Gianfreda A. and Grossi L., (2009), *Zonal Price Analysis of the Italian Wholesale Electricity Market*, IEEE CNF Conference Proceedings of the *European Energy Markets Conference* (EEM09), available on IEEE Xplore

- [21] Giulietti M., Grossi L. and Waterson M. (2010), *Price transmission in the UK electricity market: was NETA beneficial?*, Energy Economics, 32, pp. 1165–1174
- [22] GME, (2002), *Istruzioni alla disciplina del mercato italiano*, available at <http://www.mercatoelettrico.org>
- [23] GME, (2003), Technical Report April 2003 - March 2004, available at <http://www.mercatoelettrico.org>
- [24] GME, (2008), Monitoring Monthly Reports 2008, available at <http://www.mercatoelettrico.org>
- [25] GME, (2008), Technical Report 2008, available at <http://www.mercatoelettrico.org>
- [26] GME, (2009), Vademecum Italian Power Exchange, available at <http://www.mercatoelettrico.org>
- [27] Goto M. and Karolyi G.A., (2002), *Understanding Electricity Price Volatility Within and Across Markets*, working paper available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=576982
- [28] Hadsell L. and Shawky H.A., (2006), *Electricity Price Volatility and the Marginal Cost of Congestion: An Empirical Study of Peak Hours on the NYISO Market, 2001-2004*, The Energy Journal, 27, 2, pp. 157–179
- [29] Haldrup N. and Nielsen M.O., (2006), *A regime switching long memory model for electricity prices*, Journal of Econometrics, 135, 1-2, pp. 349–376
- [30] Haldrup N., Nielsen F.S. and Nielsen M.O., (2010), *A vector autoregressive model for electricity prices subject to long memory and regime switching*, Energy Economics, 32(5), pp. 1044–1058
- [31] Helman U., (2006), *Market power monitoring and mitigation in the US wholesale power markets*, Energy, 31, pp. 877-904

- [32] Hellmer S. and Warell L., (2009), *On the evaluation of market power and market dominance – The Nordic electricity market*, Energy Policy, 37, pp. 3235–3241
- [33] Huisman R. and Mahieu R., (2003), *Regime jumps in electricity prices*, Energy Economics, 25, 5, pp. 425–434
- [34] IEA PVPS, International Energy Agency – Photovoltaic Power Systems Programme, (2008), *Trends in Photovoltaic applications*, report available at http://www.iea-pvps.org/products/download/rep1/_17.pdf
- [35] Johnson B. and Barz G., (1999), *Selecting Stochastic Processes For Modelling Electricity Prices*, Energy Modelling and the Management of Uncertainty, Risk Publications
- [36] Kanamura T., Ohashi K., (2007), *A structural model for electricity prices with spikes: measurement of spike risk and optimal policies for hydropower plant operation*, Energy Economics, 29 (5), pp. 1010–1032
- [37] Karakatsani N. and Bunn D.W., (2008), *Forecasting electricity prices: The impact of fundamentals and time-varying coefficients*, International Journal of Forecasting, 24, 4, pp. 764–785
- [38] Knittel C.R. and Roberts M., (2005), *An empirical examination of restructured electricity prices*, Energy Economics, 27, 5, pp. 791–817
- [39] Koopman S. J., Ooms M., and Carnero M.A., (2007), *Periodic Seasonal Reg-ARFIMA-GARCH Models for Daily Electricity Spot Prices*, Journal of the American Statistical Association, 102, 477, pp. 16–27
- [40] Kupper G., Delarue E., Delvaux B., Meeus L., Bekaert D., Willems B., Proost S., Dhaeseleer W., Deketelaere K. and Belmans R., (2009), *Does More International Transmission Capacity Increase Competition in the Belgian Electricity Market?*, The Electricity Journal, 22, 1, pp. 21–36

- [41] Lisea W., Hobbs B.F. and Hers S., (2008), *Market power in the European electricity market – The impacts of dry weather and additional transmission capacity*, Energy Policy, 36, pp. 1331-1343
- [42] Manuhutu C. and Owen A.D., (2010), *Gas-on-gas competition in Shanghai*, Energy Policy, , 38 (5), 2101–2106.
- [43] Misiolek A., Trueck S. and Weron R., (2006) , *Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models*, Studies in Nonlinear Dynamics & Econometrics, 10, 3, article 2
- [44] Nogales F.J., Contreras J., Conejo A.J. and Espinola R., (2002) *Forecasting Next-Day Electricity Prices by Time Series Models*, IEEE Transactions on Power Systems, 17, 2, May, pp. 342–348
- [45] Mount T.D., Ning Y. and Cai X., (2006), *Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters*, Energy Economics, 28, pp. 62–80
- [46] Murry D. and Zhu Z., (2008), *Asymmetric price responses, market integration and market power: A study of the U.S. natural gas market*, Energy Economics, 30, pp. 748-765
- [47] Olbermann, B.P., Lopes, S.R.C. and Reisen, V.A. (2006), *Invariance of the first difference in ARFIMA models*, Computational Statistics, 21, pp.445-461
- [48] Petrella A. and Sapio S., (2009), *A time series analysis of day-ahead prices on the Italian Power Exchange*, IEEE CNF Proceedings of the *European Energy Markets Conference (EEM09)*, available on IEEE Xplore
- [49] Rahimi A.F. and Sheffrin A.Y., (2003), *Effective market monitoring in deregulated electricity markets.*, IEEE Transactions on Power Systems, 18, (2), pp. 486–493
- [50] Skantze P. and Illic M., (2000), *The Joint Dynamics of Electricity Spot and Forward Markets: Implications of Formulating Dynamic Hedging Strategies*, Energy Laboratory Report No. MIT-EL 00-005

- [51] Tamaschkea R., Docwrab G. and Stillman R., (2005), *Measuring market power in electricity generation: A long-term perspective using a programming model*, Energy Economics, 27, pp. 317-335
- [52] Taylor J.W. and McSharpy P.E (2007), Short-term load forecasting methods: An evaluation based on European data *IEEE Transactions on Power Systems*, 22, 2213–2219.
- [53] Juan R. Trapero J.R. and Pedregal D.J., (2009), *Frequency domain methods applied to forecasting electricity markets*, Energy Economics, 31, pp. 727–735
- [54] Weron R., (2006), *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley, New York
- [55] Weron R., Bierbrauerb M. and Truck S., (2004), *Modeling electricity prices: jump diffusion and regime switching*, Physica A, 336, pp. 39–48
- [56] Weron R. and Misiorek A., (2008), *Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models*, International Journal of Forecasting, 24, pp. 744–763
- [57] Zarnikau J. and Lam A., (1998), *Market Concentration and Market Power Among Electricity Generators in Texas*, The Electricity Journal, November, pp.60–64